CENTRE DE RECHERCHE
SUR LES TRANSPORTS
(C.R.T.)
CENTRE FOR RESEARCH
ON TRANSPORTATION

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE

CENTRE DE RECHERCHE SUR LES TRANSPORTS
UNIVERSITÉ DE MONTRÉAL
C.P. 6128, SUCCURSALE A
MONTRÉAL
CANADA H3C 3J7

TÉLÉPHONE : (514) 343-7575
EMAIL : crt@crt.umontreal.ca
TÉLÉCOPIEUR (FAX) : (514) 343-7121

92-04611/

CRT-807

A Coordinate Descent Method for the Bilevel O-D Matrix Adjustement Problem

by
Michael Florian
Yang Chen

Centre de recherche sur les transports — Publication #807
Département d'informatique et de recherche opérationnelle — Publication #802

February 1992

Dépôt légal – 2e trimestre 1992

ABSTRACT

We present a convergent descent type algorithm for the problem of adjusting an O-D matrix by using observed flows in congested networks. The problem is formulated as a bi-level optimization problem and, it is shown, in the context of the general bilevel programming problem (BLPP), that it can be reformulated as a smooth single level optimization problem. Under the general framework of the bilevel programming developed in Chen and Florian (1991), some properties of the O-D matrix adjusting problem are derived and a coordinate descent method is developed. The algorithm takes full advantage of the network structure of the second level problem and does not require many objective function evaluations. Computational results for some numerical examples are reported.

Keywords: bilevel programming, O-D matrix adjustment, equilibrium assignment, descent algorithms.

* * * * * * * * *

RÉSUMÉ

Nous présentons un algorithme convergent de descente pour le problème d'ajustement d'une matrice origine-destination à partir des comptages de trafic, sur un réseau assujetti à la congestion. Le problème est formulé comme un problème d'optimisation à deux niveaux. Nous démontrons, dans le cadre général de problèmes de programmation à deux niveaux, que le problème peut être reformulé comme un problème différentiable à un niveau. En appliquant les résultats de Chen et Florian (1991), quelques propriétés spécifiques de ce problème sont dérivées et un algorithme de descente est développé. L'algorithme exploite la structure de réseau du problème de deuxième niveau et ne requiert pas de nombreuses évaluations de la fonction économique. Des résultats numériques sont présentés également.

Mots-clés: programmation à deux niveaux, affectation d'équilibre, ajustement de la matrice O-D, algorithme de descente.

1 Introduction

The problem of estimating or adjusting an origin-destination (O-D) matrix from observed traffic flows has attracted the attention of many researchers during the past fifteen years. A large variety of models and methods have been proposed , (Nguyen (1977, 1984)), which can be classified according to the way that the observed data are used to formulate the model and the way in which the O-D matrix is assigned to the paths of the network. Van Zuylen and Willumsen (1980) propose a model for networks that have constant link travel times and extended by Fisk (1988,1989) to flow dependent link travel times. In this approach, the O-D matrix is estimated by requiring that the resulting link flows reproduce the traffic counts on a subset of links or a function of these counts. As pointed out by Fisk, the model with flow dependent link travel times leads to formulations of optimization problem with constraints which are variational inequalities or particularly, optimization problems, which lead to bilevel programming problems (BLPP). Some attempts have been made to develop practical methods to solve such problem. In the following we refer to this problem as the matrix adjustment problem, MAP.

Spiess (1990) proposes a gradient approach, based on the assumption that the path flow proportions are locally constant, and implements it by using the macro commands of the EMME/2 (INRO, 1990) software package. This approach has been applied successfully to several large scale problems originating from practice. Florian and Chen (1991) study a formulation similar to that of Spiess (1990) and convert it to a single level problem by applying KKT conditions to the second level problem. An exact penalty function approach with a trust region strategy is adapted to solve the bilevel problem. It should be pointed out

that the KKT conditions involve the addition of a large number of variables and complicated constraints to the first level problem. The resulting problems may be very large and require significant computational effort.

In general, bilevel programming problems are difficult to solve because of their inherent nonconvexity and nondifferentiability. In addition, the evaluation of the objective function and the gradient vector of the objective function at a given solution point, always require the solution to the second level problem. It is evident that numerous evaluations of the objective function are a burden for any solution method for the BLPP. Another intractability of the MAP is that, on one hand, the explicit form of the path proportion is needed in the problem formulation, on the other hand, for large scale problems it is almost impossible to work in the space of path flows directly. Therefore, it is both a theoretical and practical challenge to devise efficient algorithms for the solution of the MAP.

In the literature on bilevel programming, descent methods have been used to solve general BLPP by several authors, for example, Kolstad and Lasdon (1990) and Friesz et al. (1990). The main components of this approach are finding the descent direction and computing an optimal step length. De Silva (1978), Friesz, Tobin and Miller (1988) and Kolstad and Lasdon (1990) employ sensitivity analysis theory of nonlinear programming to calculate the derivative of the implicit function, which is defined by the functional relation between the first and second level variables, in order to determine a descent direction for the problem. Chen and Florian (1991) develop a more general approach for obtaining the descent direction by providing a set of formulae to compute the generalized gradient of the marginal function, defined to be the infimum function of the second level problem. The more difficult aspect of

a descent algorithms is to determine the step length. Heuristic methods are adapted, among others, by Friesz et al. (1990), who use a predetermined step length in order to avoid many objective function evaluations; the convergence properties of this class of methods are not guaranteed.

The aim of this work is to present a practical convergent descent type algorithm to solve the MAP when the flows on the network are determined by the user optimal principle (Wardrop, 1952). The paper is organized as follows. After this introduction, the bilevel programming formulation of the MAP problem is presented and the existence proof of solutions to the MAP is given. It is shown that, in a general framework of the BLPP's, this formulation is equivalent to a smooth single level optimization problem. Then, based on theory developed for the general nonlinear BLPP, the first-order necessary conditions for optimality of the MAP and some important properties are discussed. In order to avoid the inherent difficulties of the MAP such as the BLPP structure and the working space problems, a coordinate descent type method is derived to deal with them . The algorithm takes full advantage of the network structure of the second level problem in the way that the algorithm can work directly on the link flow space and does not require many objective function evaluations. The local convergence of the algorithm is demonstrated under certain conditions and computational results, obtained by a macro which uses the EMME/2 software package are reported. A brief conclusion ends the paper.

2 Analysis of the MAP

Let $R = (N, A)$ be a transportation network, where N is the set of nodes and $a, a \in A$ are the set of links. Denote by $g = (g_i,\ i \in I)$ the set of demands for O-D pairs, where I is the index set of all O-D pairs, and $s(v)$ the cost function defined on link flows v.

Smith (1979) shows that the mathematical model which corresponds to the user optimal path choice hypothesis reduces to the variational inequality (VI)

$$s(v^*)(v^* - v) \leq 0, \qquad \forall v \in \Omega(g) \tag{1}$$

where $\Omega(g)$ is the set of feasible link flows for a given demand g and $v^* \in \Omega(g)$ is the user optimal solution, usually referred to as the network equilibrium flows.

The MAP seeks to determine an estimate of the O-D demand g such that, the link observed volumes and possibly "a prior" information about the O-D matrix are used while the predicted link flows satisfy the user optimal principle.

The mathematical formulation of the model may be stated as follows:

$$(P1) \qquad \min_{g \leq g^{max}} F(g, v) = \gamma_1 F_1(g) + \gamma_2 F_2(v) \tag{2}$$

$$s.t. \quad s(v^*)(v^* - v) \leq 0, \quad \forall v \in \Omega(g) \tag{3}$$

where $F_1(g)$ is any suitable distance metric between the estimated trip demand and a reference trip demand matrix, and $F_2(v)$ is some suitable distance metric between the observed traffic count and the desired equilibrium flow pattern. The formulation above represents an

optimization problem with variational inequality constraints, which may be considered as a generalization of bilevel programming problem.

In order to address this problem, we make the following assumptions:

1. The demand vector is bounded, i.e. there exists g^{max} such that $0 \leq g \leq g^{max}$;
2. The cost function $s(v)$ is continuous, strictly monotone and symmetric.

We can show the existence of solutions to the problem (P1) under these assumptions.

Theorem 1 *The problem (P1) has at least one solution.*

Proof. It is known that the variational inequality $VI(1)$ has an equivalent optimization problem formulation when $s(v)$ is symmetric

$$\min f(v) = \sum_{a \in A} \int_0^{v_a} s_a(x)dx \tag{4}$$

$$s.t. \quad v \in \Omega(g). \tag{5}$$

The constraint set $\Omega(g)$ is the set of the link flows satisfying the conservation of flow and nonnegativity. By the strictly monotonicity of $s(v)$ and the compactness of $\Omega(g)$ for g fixed, $VI(1)$ has a unique solution for every given g. This establishes a functional relationship between g and v, which is denoted by $v(g)$ and is continuous on g by the sensitivity analysis theory in nonlinear programming.

Now we consider the constraint set of the problem (P1)

$$\Omega = \{(g, v) \in \mathbf{R}^{|I|+|A|} \mid v = \arg\min_{v \in \Omega(g)} f(v), \ \ g \leq g^{max}\}. \tag{6}$$

It is clear that Ω is bounded. Let (g^j, v^j), $j \in J$ be a convergent subsequence such that $v^j = \arg\min_{v \in \Omega(g^j)} f(v)$ for all $j \in J$ and (g^*, v^*) be the limit point. Let

$$\hat{v} = \arg \min_{v \in \Omega(g^*)} f(v).$$

If $\hat{v} \neq v^*$, we take $\epsilon = \frac{1}{\|\hat{v} - v^*\|} > 0$. For any $\delta \geq 0$, we can always find $g^j, j \in J$ such that

$$\|g^j - g^*\| \leq \delta, \quad \|v^j - \hat{v}\| > \epsilon,$$

which contradicts the continuity of $v(\cdot)$. Thus $(g^*, v^*) \in \Omega$ and Ω is closed. Therefore, Ω is compact. The existence of solution to the problem (P1) follows. □

In the following, we specialize the analysis for the case when the functions $F_1(g)$ and $F_2(v)$ are quadratic functions and $\gamma_1 = \gamma_2 = \frac{1}{2}$. We discuss the applicability of the approach to other suitable metrics in the sequel of the paper. With this choice, the MAP may be stated as the BLPP:

$$(P2) \qquad \min_{g \leq g^{max}} F(g, v) = \frac{1}{2}[\sum_{i \in I}(g_i - \hat{g}_i)^2 + \sum_{a \in \hat{A}}(v_a - \hat{v_a})^2] \tag{7}$$

where v solves the following problem

$$\min f(v, g) = \sum_{a \in A} \int_0^{v_a} s_a(t)dt \tag{8}$$

$$s.t. \quad v \in \Omega(g). \tag{9}$$

We shall refer to the solution of (8) and (9) as $assign(g)$.

A classical approach to deal with BLPP which have convex second level optimization problems is to derive the KKT conditions of the second level problem and add them to the

set of constraints for the first level problem. One obtains then a single level nonconvex optimization problem (Bard and Falk (1982), Bialas and Karwan (1984) Candler and Townsley (1982) and Judice and Faustino (1988)). However, as mentioned earlier, the KKT formulation may require the addition of a very large number of variables and nonlinear constraints (e.g. complementarity conditions). In particular, when the inner problem is a network equilibrium problem, the number of variables and constraints for a medium size problem is at least in the hundreds of thousands. Thus, alternative approaches to solve the problem are of interest. In order to find more efficient algorithms for the BLPP problem, Chen and Florian (1991) propose a general framework of BLPP based on the marginal function related to the second level problem and establish necessary conditions for optimality of the BLPP. By applying these results, the MAP may be reformulated as a single level optimization problem stated in the following proposition:

Proposition 1 *The (MAP) (P1) is equivalent to the following problem*

$$(P3) \qquad \min F(g,v) = \frac{1}{2}[\sum_{i\in I}(g_i - \hat{g}_i)^2 + \sum_{a\in\hat{A}}(v_a - \hat{v}_a)^2]$$

$$s.t. \quad \sum_{a\in A}\int_0^{v_a} s_a(t)dt - Z(g) = 0, \tag{10}$$

$$v \in \Omega(g) \tag{11}$$

$$g \leq g^{max} \tag{12}$$

where the marginal function $Z(g) : \mathbf{R}^{|I|} \longrightarrow \mathbf{R}$ *is defined by*

$$Z(g) = \inf\{\sum_{a\in A}\int_0^{v_a} s_a(t)dt \mid v \in \Omega(g)\}. \tag{13}$$

□

It is relevant to note that the infimum in the formulation of $Z(g)$ can be replaced by minimum due to the compactness of $\Omega(g)$. Generally, the marginal functions are no longer differentiable for nonliner second level problems, but they usually possess some sorts of (generalized) differentiabilities under different conditions. For the MAP, fortunately, the *Gâteaux – differentiability* of the marginal function $Z(g)$ can be guaranteed due to the special structure of the problem itself.

Theorem 2 *If the cost function $s(v)$ is strictly monotone, then the marginal function $Z(g)$ is Gâteaux – differentiable. The gradient of the marginal function is*

$$\nabla Z(g) = \hat{s}(g) + u(g), \tag{14}$$

where $u(g)$ is the optimal Lagrange multiplier corresponding to the second level problem for a given demand g and $\hat{s}(g)$ is given by

$$(\sum_{a \in A} \sum_{k \in K_i} \delta_{ak} p_k s_a(v_a) | i \in I),$$

p_k are path proportions for all path k. Moreover, $u(g)$ is a monotone and continuous function of g.

The continuity and monotonicity of $u(\cdot)$ have been proved by Hall (1978) and Fisk and Nguyen (1981) respectively. We give an alternative proof of continuity of $u(\cdot)$. The monotonicity of $u(\cdot)$ may also be derived from a more general result on the convexity of the marginal function in Chen and Florian (1992).

Proof. We apply Corollary 2 in Chen and Florian (1991) to the MAP. Because of the strict monotonicity of the cost function, the equilibrium assignment problem (the second level) has a unique solution for every fixed demand g. Notice that the multiplier $u(g)$ corresponds the vector of the lengths of the shortest paths between all O-D pairs of the traffic network, based on the current link cost $s(v)$. Thus $u(g)$ is unique. Therefore, the gradient with respect to g of the Lagrangien is

$$\nabla Z(g) = \nabla_g(\sum_{a\in A}\int_0^{v_a} s_a(t)dt) + u(g) + 0\cdot\mu = \hat{s}(g) + u(g),$$

where μ is the multiplier vector corresponding to the constraints $h \geq 0$.

By the theory of sensitivity analysis, the marginal function $Z(\cdot)$ and the optimal solution $h^*(\cdot)$ of the equilibrium assignment problem are continuous on g, where $h^*(g)$ is the equilibrium path flow for the given g. This implies that the cost function $s(v)$ is also a continuous function of g. Thus, for any $\epsilon > 0$, there exist $\delta > 0$ and a sufficiently large number $M > 0$ such that

$$|h_k^*(g) - h_k^*(g')| < \epsilon/M, \tag{15}$$

for all $k \in K_i$, $i \in I$, and

$$|s_a(v^*(g)) - s_a(v^*(g'))| < \epsilon/M, \tag{16}$$

for all $a \in A$, provided for all g' in the ball centered by g

$$\|g - g'\| < \delta.$$

For any O-D pair $i \in I$, there can only be two cases:

Case 1. $h_k^*(g) = 0$ for all paths $k \in K_i$. Consequently, the cost of this O-D pair $u_i(g)$ is zero. It is easily to see that $h_k^*(g') = 0$ also holds because of (15) and the arbitrariness of ϵ.

Therefore

$$|u_i(g')| = |u_i(g) - u_i(g')| = 0. \tag{17}$$

Case 2. If there exists a path $k' \in K_i$ such that $h^*_{k'}(g) > 0$, then

$$u_i(g) = \sum_{a \in A} \delta_{ak'} s_a(v^*(g)).$$

For the same reason in Case 1, $h^*_{k'}(g') > 0$ must be true. Thus,

$$u_i(g') = \sum_{a \in A} \delta_{ak'} s_a(v^*(g')).$$

Therefore,

$$\begin{aligned} |u_i(g) - u_i(g')| &= |\sum_{a \in A} \delta_{ak}(s_a(v^*(g)) - s_a(v^*(g')))| \\ &< |A|\epsilon/M < \epsilon. \end{aligned} \tag{18}$$

We conclude that $u(\cdot)$ is continuous on g by (17) and (18). □

In order to obtain further insight in the structure of the MAP, it is necessary to reformulate the problem in terms of the path proportions p. The MAP is reformulated as follows:

$$(P4) \qquad \min F(g,h) = \frac{1}{2}[\sum_{i \in I}(g_i - \hat{g}_i)^2 + \sum_{a \in A}(v_a - \hat{v}_a)^2]$$

$$s.t. \quad \sum_{a \in A} \int_0^{v_a} s_a(t)dt - Z(g) = 0, \tag{19}$$

$$\sum_{k \in K_i} h_k - g_i = 0, \quad i \in I \tag{20}$$

$$h_k = p_k g_i, \quad k \in K_i,\ i \in I \tag{21}$$

$$g \le g^{max}, \tag{22}$$

$$h \ge 0, \tag{23}$$

where p_k is the proportion of the demand g_i which is assigned to path k, $k \in K_i$ and $v_a = \sum_{i \in I} \sum_{k \in K_i} \delta_{ak} h_k$ for all $a \in A$ and δ_{ak} is defined as

$$\delta_{ak} = \begin{cases} 1 & \text{if path k passes through link a} \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

The constraint (21) is important in the formulation of the MAP, since the path proportions are not constants in the network equilibrium problem for congested traffic networks. Substituting h_k by $p_k g_i$, one obtains

$$(P5) \qquad \min F(g,h) = \frac{1}{2}[\sum_{i \in I}(g_i - \hat{g}_i)^2 + \sum_{a \in \hat{A}}(v_a - \hat{v}_a)^2]$$

$$s.t. \quad \sum_{a \in A} \int_0^{v_a} s_a(t)dt - Z(g) = 0, \tag{25}$$

$$\sum_{k \in K_i} p_k g_i - g_i = 0 \qquad i \in I \tag{26}$$

$$g \leq g^{max}, \tag{27}$$

$$g \geq 0, \tag{28}$$

$$h \geq 0. \tag{29}$$

It is clear that, from a practical viewpoint, the demand can be assumed to be positive. Then (P5) simplifies to:

$$(P6) \qquad \min F(g,p) = \frac{1}{2}[\sum_{i \in I}(g_i - \hat{g}_i)^2 + \sum_{a \in \hat{A}}(v_a - \hat{v}_a)^2]$$

$$s.t. \quad \sum_{a \in A} \int_0^{v_a} s_a(t)dt - Z(g) = 0, \tag{30}$$

$$\sum_{k \in K_i} p_k - 1 = 0 \qquad i \in I^+ \tag{31}$$

$$g \leq g^{max}, \tag{32}$$

$$g \geq 0, \tag{33}$$

$$p \geq 0, \tag{34}$$

where

$$I^+ = \{i \mid g_i > 0\}.$$

Now we are ready to establish the necessary conditions for an optimal solution of the MAP.

Theorem 3 *If (g^*, p^*) solves the problem (P1) and a constraint qualification holds at (g^*, p^*) corresponding to (P6), then there exist $\alpha \in \mathbf{R}$, $u \in \mathbf{R}^{|I|}$, $\eta \in \mathbf{R}^{|I|}$, $\xi \in \mathbf{R}^{|I|}$ and $\phi \in \mathbf{R}^{|I|}$ such that*

$$p_k^*(\sum_{a \in A} \delta_{ak} s_a(v_a^*) - u_i) = 0, \tag{35}$$

$$\sum_{a \in A} \delta_{ak} s_a(v_a^*) - u_i \geq 0, \tag{36}$$

$$p_k^*(\sum_{a \in A} \delta_{ak} c_a^* \sum_{i \in I} g_i^* - \eta_i) = 0, \tag{37}$$

$$\sum_{a \in A} \delta_{ak} c_a^* \sum_{i \in I} g_i^* - \eta_i \geq 0, \tag{38}$$

$$(g_i^* - g_i^{max})(g_i^* - \hat{g}_i - \alpha u_i + \sum_{a \in A} \sum_{k \in K_i} \delta_{ak} c_a^* - \xi_i) = 0, \tag{39}$$

$$\alpha u_i - (g_i^* - \hat{g}_i) - \sum_{a \in A} \sum_{k \in K_i} \delta_{ak} c_a^* + \xi_i \geq 0, \tag{40}$$

$$g_i^*(g_i^* - \hat{g}_i - \alpha u_i + \sum_{a \in A} \sum_{k \in K_i} \delta_{ak} c_a^* + \phi_i) = 0, \tag{41}$$

for all $k \in K_i$ and $i \in I$, where

$$c_a^* = \begin{cases} v_a^* - \hat{v}_a + \alpha s_a(v_a^*) & a \in \hat{A} \\ \alpha s_a(v_a^*) & a \in A - \hat{A} \end{cases} \tag{42}$$

Proof. By applying Corollary 3 in Chen and Florian (1991) to the problem (P6), we obtain the conditions (37) – (41). Notice that (30) and feasibility conditions imply that p^* is an equilibrium flow proportion for $g = g^*$, which is equivalent to (35) and (36). Then the theorem is proved due to the uniqueness of the optimal multiplier of the second level problem. □

The conditions (35)-(41) indicate how an optimal solution to the MAP behaves. (35)-(36) imply that the optimal solution must observe Wardrop's principle, that means it must be an equilibrium flow pattern. (37)-(38) may also be viewed as an expression of Wardrop's principle on a modified network, which takes into account both the first and the second level problems, where c_a^*, $a \in A$ are the link costs and η_i, $i \in I$ can be interpreted as the shortest path lengths on the modified network.

3 The Method

In this section, we propose a coordinate descent type method to solve problem (P6). As mentioned in the introduction, the difficulties involved in this problem are twofold: any possible solution method must cope with the two level structure of the problem and, in addition, the method must use the path proportion explicitly. The algorithm presented below satisfies both of these requirements.

Consider a feasible solution (g^l, p^l) of (P6); it gives rise to two subproblems, which are obtained by fixing the variables g^l and p^l in turn. The constraint relaxed subproblem for $p = p^l$ is defined as follows:

$$(P6)_{p^l} \qquad \min F(g,p^l) = \frac{1}{2}[\sum_{i\in I}(g_i-\hat{g}_i)^2 + \sum_{a\in\hat{A}}(v_a(g)-\hat{v}_a)^2]$$
$$s.t. \qquad 0 \le g \le g^{max}.$$

By performing a descent step on the objective function $F(g,p^l)$, a new g^{l+1} is generated. Then the restricted form of (P6) for $g = g^{l+1}$ is given by

$$(P6)_{g^{l+1}} \qquad \min F(g^{l+1},h) = \frac{1}{2}[\sum_{i\in I}(g_i^{l+1}-\hat{g}_i)^2 + \sum_{a\in\hat{A}}(v_a(p)-\hat{v}_a)^2]$$
$$s.t. \quad \sum_{a\in A}\int_0^{v_a} s_a(t)dt - Z(g^{l+1}) = 0,$$
$$\sum_{k\in K_i} p_k - 1 = 0, \qquad i \in I^+$$
$$p \ge 0.$$

The algorithm is designed to work alternatively on these two subproblems and is described below.

The Algorithm

Step 1. Choose an initial demand matrix g^1; solve the network equilibrium problem with $g = g^1$. p^1 is obtained implicitly by the solution $v(g^1,p^1)$. Set l:=1.

Step 2. g^l and p^l are known. Solve the direction finding subproblem

$$\min \sum_{i\in I} w_i^l x_i \tag{43}$$
$$s.t. \quad 0 \le x_i \le g_i^{max}, \quad i \in I^+, \tag{44}$$

where, for all $i \in I$,

$$w_i^l = \frac{\partial F(g, p^l)}{\partial g_i} \mid_{g_i^l} \tag{45}$$

$$= g_i^l - \hat{g}_i + \sum_{a \in \hat{A}} \sum_{k \in K_i} \delta_{ak}(v_a^l - \hat{v}_a) p_k^l. \tag{46}$$

for the problem $(P6)_{p^l}$. The descent direction is given by $d_g^l = x^l - g^l$.

Step 3. If $d_g^l = 0$, terminate the algorithm; otherwise, let

$$g^{l+1} = g^l + \lambda^l d_g^l, \tag{47}$$

where λ^l is determined by

$$F(g^l + \lambda d_g, p^l) - F(g^l, p^l) \leq \frac{\lambda}{2} \nabla F(g^l, p^l) d_g^l, \tag{48}$$

$$0 \leq \lambda \leq 1, \tag{49}$$

and continue to Step 4.

Step 4. Solve the subproblem $(P6)_{g^{l+1}}$. Denote the solution by p^{l+1}, go to Step 2.

It is important to note that to solve $(P6_{g^{l+1}})$ in Step 4 is equivalent to solve an equilibrium assignment problem $assign(g^{l+1})$, since the equilibrium link flow v^{l+1} is unique for a given g^{l+1}, and so is the objective function value, no matter what p^{l+1} is. It will be shown that, the algorithm requires solving one equilibrium assignment problem for Step 4 and at most two additional assignment problems for Step 2 and Step 3 in each iteration. This is stated formally in the following lemmas.

Lemma 1 *The solution to the direction finding subproblem in Step 2 at each iteration is computed by*

$$x_i^l = \begin{cases} g_i^{max} & if\ w_i^l < 0 \\ g_i^l & if\ w_i^l = 0 \\ 0 & if\ w_i^l > 0 \end{cases}$$

□

It should be pointed out that the quantity w^l can be determined by using the *additional options* in EMME/2 software package after solving the equilibrium assignment problem as same as in (Spiess, 1990).

Lemma 2 *The step length λ^l at each iteration can be determined explicitly by*

$$0 \le \lambda_l \le \min(1, \frac{\sum_{i\in I}(g_i^l - \hat{g}_i)(g_i^l - x_i^l) + \sum_{a\in\hat{A}}(v_a^l - \hat{v}_a)(v_a^l - v_a(x^l))}{\sum_{i\in I}(g_i^l - x_i^l)^2 + \sum_{a\in\hat{A}}(v_a^l - v_a(x^l))^2})$$

where $v_a(x^l) = \sum_{i\in I}\sum_{k\in K_i}\delta_{ak}p_k^l x_i^l$ for $a \in \hat{A}$.

Proof. By the definition of F,

$$\begin{aligned} & F(g^{l+1}, p^l) - F(g^l, pl) \\ &= \frac{1}{2}\{\sum_{i\in I}[(g_i^l + \lambda d_{g_i}^l - \hat{g}_i)^2 - (g_i^l - \hat{g}_i)^2] + \sum_{a\in\hat{A}}[(\sum_{i\in I}\sum_{k\in K_i}\delta_{ak}p_k^l(g_i^l + \lambda d_{g_i}^l) - \hat{v}_a)^2 - (v_a^l - \hat{v}_a)^2]\} \\ &= \frac{\lambda}{2}\{\sum_{i\in I}[\lambda(g_i^l - x_i^l)^2 - 2(g_i^l - x_i^l)(g_i^l - \hat{g}_i)] + \sum_{a\in\hat{A}}[\lambda(v_a^l - v_a(x^l))^2 - 2(v_a^l - v_a(x^l))(v_a^l - \hat{v}_a)]\}. \end{aligned}$$

Thus, the inequality (48) implies that

$$\begin{aligned} & [\sum_{i\in I}(g_i^l - x_i^l)^2 + \sum_{a\in\hat{A}}(v_a^l - v_a(x^l))^2]\lambda \\ &\le \sum_{i\in I} w_i^l(x_i^l - g_i^l) + 2\sum_{i\in I}(g_i^l - x_i^l)(g_i^l - \hat{g}_i) + 2\sum_{a\in\hat{A}}(v_a^l - v_a(x^l))(v_a^l - \hat{v}_a) \\ &= \sum_{i\in I}(g_i^l - \hat{g}_i)(g_i^l - x_i^l) + \sum_{a\in\hat{A}}(v_a^l - \hat{v}_a)(v_a^l - v_a(x^l) \end{aligned}$$

Therefore,

$$0 \leq \lambda \leq \min(1, \frac{\sum_{i \in I}(g_i^l - \hat{g}_i)(g_i^l - x_i^l) + \sum_{a \in \hat{A}}(v_a^l - \hat{v}_a)(v_a^l - v_a(x^l))}{\sum_{i \in I}(g_i^l - x_i^l)^2 + \sum_{a \in \hat{A}}(v_a^l - v_a(x^l))^2})$$

□

Lemma 3 *The sequence (g^l, p^l) generated by the algorithm is a sequence of feasible point for the problem (P6).*

Proof. At any iteration l, g^l is given by

$$\begin{aligned} g^l &= g^{l-1} + \lambda^{l-1}(x^{l-1} - g^{l-1}) \\ &= (1 - \lambda^{l-1})g^{l-1} + \lambda l - 1 x^{l-1}, \end{aligned}$$

which is in fact

$$g_i^l = \begin{cases} (1 - \lambda^{l-1})g_i^{l-1} + \lambda^{l-1} g_i^{max} & \text{if } w_i^l < 0 \\ g_i^{l-1} & \text{if } w_i^l = 0 \\ (1 - \lambda^{l-1})g_i^{l-1} & \text{if } w_i^l > 0 \end{cases}$$

By Lemma 2, we conclude that $0 \leq g^l \leq g^{max}$. It is also clear to see that, from Step 4 of the algorithm, (g^l, p^l) satisfies the constraints (30), (31) and (34). □

Now, we can proceed to prove the convergence of the algorithm under certain conditions.

Theorem 4 *If the algorithm terminates at (g^l, p^l) in a finite number of iterations, then (g^l, p^l) satisfies the necessary optimality conditions stated in Theorem 3. Otherwise, the infinite sequence $\{(g^l, p^l)\}$ generated by the algorithm converges to a local minimum.*

Proof. We discuss the two cases respectively.

1. By Lemma 3, (g^l, p^l) is feasible for (P6), which means that p^l are equilibrium path flow proportion that correspond to g^l. Thus the Wardrop's principle (35) and (36) is satisfied. The descent direction at (g^l, p^l) must be zero, which implies that $w^l = 0$. Notice that

$$\begin{aligned} w_i^l &= g_i^l - \hat{g}_i + \sum_{a \in \hat{A}} \sum_{k \in K_i} \delta_{ak} p_k^l (v_a^l - \hat{v}_a) \\ &= g_i^l - \hat{g}_i + \sum_{a \in \hat{A}} \sum_{k \in K_i} \delta_{ak} p^l c_a^l - \alpha u_i - \xi_i, \qquad i \in I \end{aligned}$$

where $\alpha = 0$, $\xi = 0$ and

$$c_a^l = \begin{cases} v_a^l - \hat{v}_a + \alpha s_a(v_a^l) & \text{if } a \in \hat{A} \\ \alpha s_a(v_a^l) & \text{if } a \in A - \hat{A} \end{cases}$$

Therefore, the conditions (39)-(41) are satisfied at (g^l, p^l) by letting $\alpha = 0$, $\xi = 0$ and $\phi = 0$. For the conditions (37)-(38), we simply let

$$\eta_i = \min_{k \in K_i} (\sum_{a \in A} \delta_{ak} c_a^l \sum_{i \in I} g_i^l), \qquad i \in I$$

2. It is sufficient to prove the latter part of the theorem by showing that the sequence $\{F(g^l, p^l)\}$ is a monotone decreasing sequence under certain conditions.

Consider the sequence $\{F(g^{l+1}, p^l)\}$. It can be seen from the algorithm that

$$F(g^{l+1}, p^l) < F(g^l, p^l)$$

for any l. Let $\epsilon^l = F(g^l, p^l) - F(g^{l+1}, p^l) > 0$. Since the optimal flow vector h is a continuous function of the demand g by the result of Theorem 2, then for any $\epsilon' > 0$, there exists $\delta' > 0$, such that

$$| h_k^{l+1} - h_k^l | < \frac{\epsilon'}{2}$$

whenever $| g_i^{l+1} - g_i^l | < \delta'$ for $k \in K_i$, $i \in I$. By the definition of h, we obtain that

$$| p_k^{l+1} g_i^{l+1} - p_k^l g_i^l | = | (p_k^{l+1} - p_k^l) + \lambda p_k^{l+1} d_{g_i}^l | < \frac{\epsilon'}{2}.$$

Thus

$$| p_k^{l+1} - p_k^l | < \epsilon', \quad k \in K_i, i \in I \tag{50}$$

if λ is small enough.

By the continuity of $F(\cdot,\cdot)$, for $\epsilon^l > 0$, there exists a $\delta(\epsilon^l) > 0$ such that

$$| F(g^{l+1}, p^{l+1}) - F(g^{l+1}, p^l) | < \epsilon^l$$

if $| p_k^{l+1} - p_k^l | < \delta(\epsilon^l)$ for all k. The existence of such a δ is guaranteed by (50). Therefore,

$$F(g^{l+1}, p^{l+1}) < F(g^{l+1}, p^l) + \epsilon^l = F(g^l, P^l).$$

□

4 A Numerical Example

In order to illustrate the performance of the algorithm we carry out a test in a small network that is taken from Nguyen and Dupuis (1983). The network has 20 links and 4 O-D pairs, which is illustrated below in Figure 1. The link cost function are defined in the form of

$$s_a(v_a) = \psi_a + \kappa_a v_a^2, \quad a \in A,$$

where ψ_a and κ_a are constants.

The test is constructed as follow. The matrix

$$\begin{pmatrix} 400 & 800 \\ 600 & 200 \end{pmatrix}.$$

is taken as a target matrix and is assigned to the network. The equilibrium flows v^* given by the target matrix are modified randomly in the interval $[v^* - 10\%,\ v^* + 10\%]$ and is taken as

the "observed" flows. [1] Table 1 gives the observed flows and the corresponding coefficients of the cost function.

Given the subset of the links with observed flows $\hat{A}$, which consists of links 2, 4, 5, 9, 14, and 19, the algorithm begins at an initial matrix

$$\begin{pmatrix} 460 & 740 \\ 749 & 51 \end{pmatrix},$$

which is obtained by perturbing the target matrix, and obtaines the predicted matrix as

$$\begin{pmatrix} 374.9 & 777.7 \\ 589.2 & 231.8 \end{pmatrix}$$

after seven iterations. If the initial matrix is taken to be the target matrix, the algorithm converges to another local solution:

$$\begin{pmatrix} 399.9 & 799.8 \\ 600.3 & 200.4 \end{pmatrix}.$$

The computational results are listed in Appendix. The fits of predicted vs. observed flows before and after adjustment are also given in Figure 2 and Figure 3 in Appendix.

5 Conclusion

In this paper, we present a successive linear approximation method to solve the MAP on congested network supported by the theory of general nonlinear bilevel programming problem. The algorithm takes full advantage of the network structure of the second level problem; it operates in the space of arc flows and may be implemented by using a widely available software package. The convergence properties of the algorithm and the relatively few evaluations of the objective function distinguish it from some other descent type methods for BLPP. It is

[1] This was discussed with J. Lundgren and O. Drissi-Kaïtouni.

worthwhile to note that the method can be extended to the MAP with variational inequality constraints without difficulty. The same approach may even be applied to general convex BLP problems which have a network structure for the lower level problem.

It remains to test the method on large scale problems. This should be relatively efficient since, at each iteration, at most three network equilibrium problems are solved. While we specialized the analysis for the choice of quadratic functions $F_1(\cdot)$ and $F_2(\cdot)$ with $\gamma_1 = \gamma_2 = \frac{1}{2}$, it is clear that any other metric, which result in a differentiable convex function at the first level of the BLPP formulation of the MAP, may be solved by this approach. Clearly, the step length computation will no longer be given by an analytical formula in general. Finally, we would like to mention that this analysis and algorithm were motivated to a large extent by the successful application of the approximate gradient of Spiess (1990) for this problem.

Acknowledgment

We would like to express our appreciates to Dr. Jan Lundgren for a very careful reading of an earlier version of this paper. The EMME/2 macro which implements this algorithm is based on a modification of the DEMADJ macro written by H. Spiess to implement his gradient method (1990).

References

[1] Bard, J.F. and J.E. Falk (1982), An Explicit Solution to the Multilevel Programming Problem, *Computers and Operations Research* 9, 77-100.

[2] Bialas, W.F. and Karwan (1984), Two-level Linear Programming, *Management Science* 30, 1004-1020.

[3] Candler, W. and R. Townsley (1982), A Linear Two-level Programming Problem, *Computers and Operations Research* 9, 59-76.

[4] Cascetta, E. and S. Nguyen (1988), A Unified Framework for Estimating or Updating Origin-Destination Matrices from Traffic Counts, *Transportation Research B* 22, 437-455

[5] Chen, Y. and Florian, M. (1991), The Nonlinear Bilevel Programming Problem: A General Formulation and Optimality Conditions, *Publication No.794, CRT, Université de Montréal, Montréal, Canada* .

[6] Chen, Y. and Florian, M. (1992), A Global Optimal Solution Method for a Class of Nonlinear Bilevel Programming Problems, *Working paper*

[7] De Silva, A.H. (1978), Sensitivity Formulas for Nonlinear Factorable Programming and Their Application to the Solution of an implicitly Defined Optimization Model of US Crude oil Production, Dissertation, George Washington University, Washington D.C. .

[8] Erlander, S., S. Nguyen and N. Stewart (1979), On the Calibration of the Combined Distribution-Assignment Model, *Transportation Research B* 13, 259-267.

[9] Fisk, C.S. (1988), On Combining Maximum Entropy Trip Matrix Estimation with User Optimal Assignment, *Transportation Research B* 22, 69-73.

[10] Fisk, C.S. (1989), Trip Matrix Estimation from Link Traffic Counts: The Congested Network Case, *Transportation Research B* 23, 331-336.

[11] Fisk, C.S. and D. Boyce (1983), A Note on Trip Matrix Estimation from Link Traffic Count Data, *Transportation Research B* 17, 245-250.

[12] Fisk, C.S. and S. Nguyen (1981), Existence and Uniqueness Properties of an Asymmetric Two-Mode Equilibrium Model, *Transportation Science* 15, 318-328.

[13] Florian, M. (1986), Nonlinear Cost Network Models in Transportation Analysis, *Mathematical Programming Study* , 167-196.

[14] Florian, M. and Chen, Y. (1991), A Bilevel Programming Approach to Estimating O-D Matrix by Traffic Counts, *Publication No. 750, CRT, Université de Montréal, Montréal, Canada.*

[15] Frank, M. and Wolfe, P. (1956), An Algorithm for Quadratic Programming, *Naval Res. Logist. Quart* 3, 95-110.

[16] Friesz, T.L., R.L. Tobin, H. Cho and N.J. Mehta (1990), Sensitivity Analysis Based Heuristic Algorithms for Mathematics Programs with Variational Inequality Constraints, *Mathematical Programming* 48, 265-284.

[17] Friesz, T.L., R.L. Tobin and T. Miller (1988), Algorithms for Spatially Competitive Network Facility-Location, *Environment and Planning B* 15, 191-203.

[18] Hall, M. A. (1978), Properties of the Equilibrium State in Transportation Networks, *Transportation Science* 12, 208-216.

[19] INRO Consultants Inc. (1990), EMME/2 Users' Manual, 735 pp, Montréal, Québec, Canada.

[20] Kolstad, C.D. and L.S. Lasdon (1990), Derivative Evaluation and Computational Experience with Large Bilevel Mathematical Programming, *Journal of Optimiza tion and Applications* 65, 485-499.

[21] LeBlanc, L. J., Morlok, E. K. and Pierskalla, W. P. (1975), An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem, *Transportation Research* 9, 309-318.

[22] Nguyen, S. (1977), Estimation an O-D Matrix from Network Data: A Network Equilibrium Approach, *Publication No. 60, CRT, Université de Montréal, Montréal, Canada*

[23] Nguyen, S. (1984), Estimating Origin-Destination Matrices from Observed Flows, In *Transportation Planning Models* (Eds. M. Florian) 363-380.

[24] Nguyen, S. and C. Dupuis (1983), An Efficient Method for Computing Traffic Equilibria in Networks with Asymmetric Transportation Costs, *Publication No. 205, CRT, Université de Montréal, Montréal, Canada.*

[25] Spiess, H. (1990), A Gradient Approach for the O-D Matrix Adjustment Problem, *Publication No.693, CRT, Université de Montréal, Montréal, Canada*

[26] Smith, M.J. (1979), Existence, Uniqueness and Stability of Traffic Equilibria, *Transportation Research B* 13, 295-304.

[27] Van Zuylen, J.H. and L.G. Willumsen (1980), The Most Likely Trip Matrix Estimated from Traffic Counts, *Transportation Research B* 14, 281-293.

[28] Wardrop, J. G. (1952), Some Theoretical Aspects of Road Traffic Research, *Proceedings of the Insititute of Civil Engineers* Part II, 325-378.

Appendix

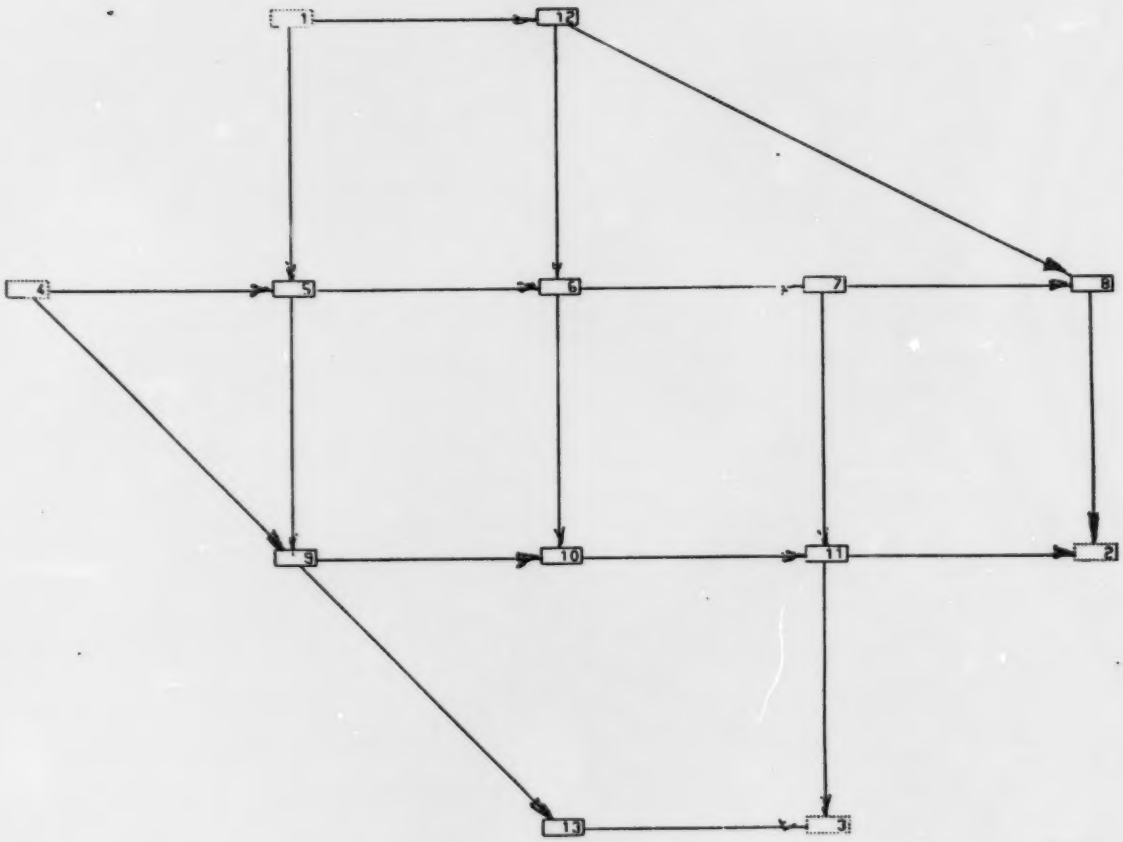

Figure 1: The test network.

links	observed flows	ψ_a	κ_a
(1,5)	600	7	1.875×10^{-5}
(1,12)	490	9	1.5×10^{-5}
(4,5)	164	9	1.5×10^{-5}
(4,9)	680	12	0.75×10^{-5}
(5,6)	476	3	1.125×10^{-5}
(5,9)	318	9	1.125×10^{-5}
(6,7)	519	5	1.875×10^{-5}
(6,10)	118	13	0.75×10^{-5}
(7,8)	149	5	1.875×10^{-5}
(7,11)	352	9	1.875×10^{-5}
(8,2)	545	9	1.875×10^{-5}
(9,10)	421	10	0.75×10^{-5}
(9,13)	532	9	0.75×10^{-5}
(10,11)	594	6	0.375×10^{-5}
(11,2)	449	9	0.75×10^{-5}
(11,3)	488	8	1.5×10^{-5}
(12,6)	150	7	0.375×10^{-5}
(12,8)	406	14	1.5×10^{-5}
(13,3)	540	11	1.5×10^{-5}

Table 1: The observed link volumes and the coefficients of the link cost functions.

links	v/d function	observed fs.	predicted fs.	obj. func. value	iter. No.
(1,12)	2	490	513.987		
(4,9)	3	680	648.308		
(5,6)	4	476	472.481		
(7,8)	6	149	188.708		
(10,11)	10	594	616.232		
(13,3)	14	540	561.045		
				3179.4907	
					7

Table 2: The predicted link flows and the objective function value.

links	v/d function	observed fs.	predicted fs.	obj. func. value	iter. No.
(1,12)	2	490	530.259		
(4,9)	3	680	630.201		
(5,6)	4	476	467.764		
(7,8)	6	149	189.019		
(10,11)	10	594	603.159		
(13,3)	14	540	563.27		
				3197.883	
					3

Table 3: The prediced link flows and the objective function value with the target matrix taken to be the initial matrix.

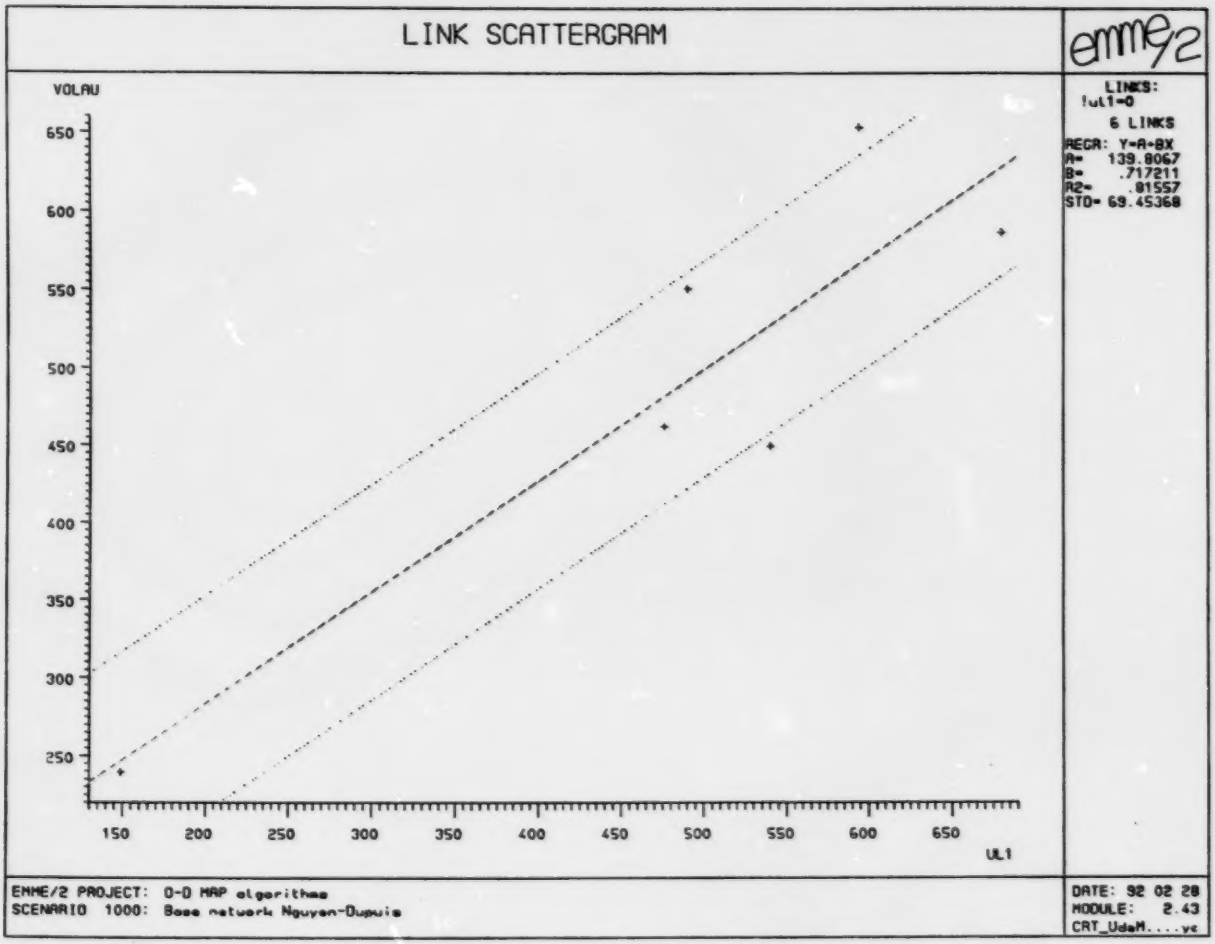

Figure 2: The fit of initial flows vs. observed flows.

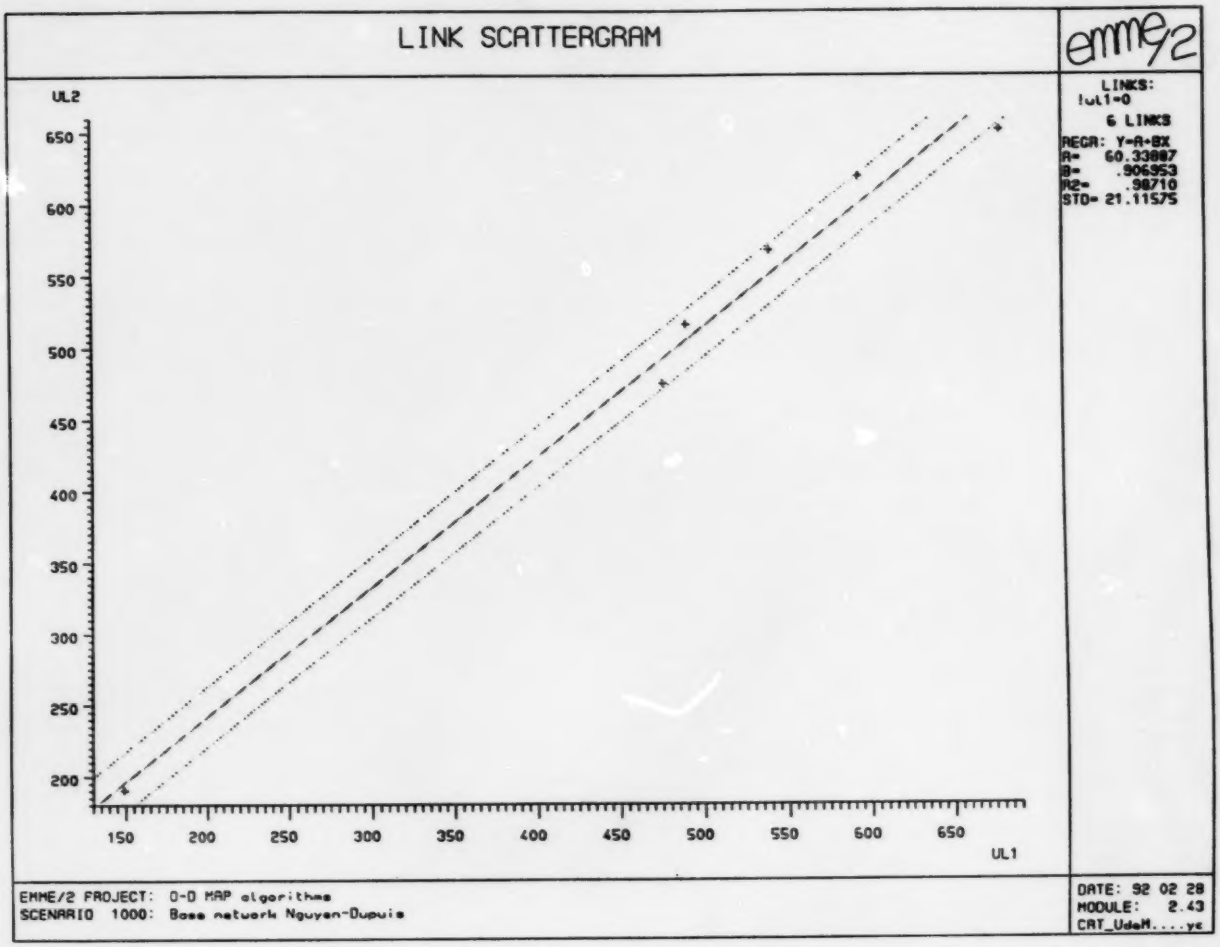

Figure 3: The fit of predicted vs. observed flows.